# LetsGrowMore Virtual Internship Program

# Intermediate level Task-3 Exploratory Data Analysis on Dataset - Terrorism

# Dinesh Chandra Gaddam

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```
df=pd.read_csv("S:\\documents\\all courses\\data_sets\\globalterrorism.csv", encoding= 'latin1')
```

In [3]:
```
df.head(5)
```

Out[3]:

| | eventid | iyear | imonth | iday | approxdate | extended | resolution | country | country_txt | region | ... | addnotes | scite1 | scite2 | scite3 | dbsour |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1.970000e+11 | 1970 | 7 | 2 | NaN | 0 | NaN | 58 | Dominican Republic | 2 | ... | NaN | NaN | NaN | NaN | PG |
| **1** | 1.970000e+11 | 1970 | 0 | 0 | NaN | 0 | NaN | 130 | Mexico | 1 | ... | NaN | NaN | NaN | NaN | PG |
| **2** | 1.970000e+11 | 1970 | 1 | 0 | NaN | 0 | NaN | 160 | Philippines | 5 | ... | NaN | NaN | NaN | NaN | PG |
| **3** | 1.970000e+11 | 1970 | 1 | 0 | NaN | 0 | NaN | 78 | Greece | 8 | ... | NaN | NaN | NaN | NaN | PG |
| **4** | 1.970000e+11 | 1970 | 1 | 0 | NaN | 0 | NaN | 101 | Japan | 4 | ... | NaN | NaN | NaN | NaN | PG |

5 rows × 135 columns

In [4]:
```
df.tail()
```

Out[4]:

| | eventid | iyear | imonth | iday | approxdate | extended | resolution | country | country_txt | region | ... | addnotes | scite1 | scit |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **181686** | 2.020000e+11 | 2017 | 12 | 31 | NaN | 0 | NaN | 182 | Somalia | 11 | ... | NaN | "Somalia: Al-Shabaab Militants Attack Army Che... | "Highligh Soma Da Mec Highligh 2 |
| **181687** | 2.020000e+11 | 2017 | 12 | 31 | NaN | 0 | NaN | 200 | Syria | 10 | ... | NaN | "Putin's 'victory' in Syria has turned into a ... | "Tv Russi soldie killed Hmeymi base |
| **181688** | 2.020000e+11 | 2017 | 12 | 31 | NaN | 0 | NaN | 160 | Philippines | 5 | ... | NaN | "Maguindanao clashes trap tribe members," Phil... | Na |
| **181689** | 2.020000e+11 | 2017 | 12 | 31 | NaN | 0 | NaN | 92 | India | 6 | ... | NaN | "Trader escapes grenade attack in Imphal," Bus... | Na |
| **181690** | 2.020000e+11 | 2017 | 12 | 31 | NaN | 0 | NaN | 160 | Philippines | 5 | ... | NaN | "Security tightened in Cotabato following IED ... | "Secur tighten in Cotaba City Manila |

5 rows × 135 columns

In [5]:

```
df.dtypes.to_frame()
```

Out[5]:

| | 0 |
|---|---|
| **eventid** | float64 |
| **iyear** | int64 |
| **imonth** | int64 |
| **iday** | int64 |
| **approxdate** | object |
| **...** | ... |
| **INT_LOG** | int64 |
| **INT_IDEO** | int64 |
| **INT_MISC** | int64 |
| **INT_ANY** | int64 |
| **related** | object |

135 rows × 1 columns

In [6]:

```
df.columns.to_frame().head(20)
```

Out[6]:

| | 0 |
|---|---|
| **eventid** | eventid |
| **iyear** | iyear |
| **imonth** | imonth |
| **iday** | iday |
| **approxdate** | approxdate |
| **extended** | extended |
| **resolution** | resolution |

| | 0 |
|---|---|
| **country** | country |
| **country_txt** | country_txt |
| **region** | region |
| **region_txt** | region_txt |
| **provstate** | provstate |
| **city** | city |
| **latitude** | latitude |
| **longitude** | longitude |
| **specificity** | specificity |
| **vicinity** | vicinity |
| **location** | location |
| **summary** | summary |
| **crit1** | crit1 |

## considering all the rows that are particularly needed

In [7]:

```
data=df[['iyear','imonth','iday','country_txt','provstate','region_txt','latitude','longitude','success','attacktype1_txt','
data.head()
```

Out[7]:

| | iyear | imonth | iday | country_txt | provstate | region_txt | latitude | longitude | success | attacktype1_txt | city | targtype1_txt | motive |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1970 | 7 | 2 | Dominican Republic | NaN | Central America & Caribbean | 18.456792 | -69.951164 | 1 | Assassination | Santo Domingo | Private Citizens & Property | NaN |
| **1** | 1970 | 0 | 0 | Mexico | Federal | North America | 19.371887 | -99.086624 | 1 | Hostage Taking (Kidnapping) | Mexico city | Government (Diplomatic) | NaN |

| | iyear | imonth | iday | country_txt | provstate | region_txt | latitude | longitude | success | attacktype1_txt | city | targtype1_txt | motive |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2** | 1970 | 1 | 0 | Philippines | Tarlac | Southeast Asia | 15.478598 | 120.599741 | 1 | Assassination | Unknown | Journalists & Media | NaN |
| **3** | 1970 | 1 | 0 | Greece | Attica | Western Europe | 37.997490 | 23.762728 | 1 | Bombing/Explosion | Athens | Government (Diplomatic) | NaN |
| **4** | 1970 | 1 | 0 | Japan | Fukouka | East Asia | 33.580412 | 130.396361 | 1 | Facility/Infrastructure Attack | Fukouka | Government (Diplomatic) | NaN |

## considering all rows that have success attempt 1 i.e successfully harmed people

In [8]:

```
dff=data[data['success']==1]
print((dff.shape))
dff.head()
```

```
(161632, 15)
```

Out[8]:

| | iyear | imonth | iday | country_txt | provstate | region_txt | latitude | longitude | success | attacktype1_txt | city | targtype1_txt | motive |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1970 | 7 | 2 | Dominican Republic | NaN | Central America & Caribbean | 18.456792 | -69.951164 | 1 | Assassination | Santo Domingo | Private Citizens & Property | NaN |
| **1** | 1970 | 0 | 0 | Mexico | Federal | North America | 19.371887 | -99.086624 | 1 | Hostage Taking (Kidnapping) | Mexico city | Government (Diplomatic) | NaN |
| **2** | 1970 | 1 | 0 | Philippines | Tarlac | Southeast Asia | 15.478598 | 120.599741 | 1 | Assassination | Unknown | Journalists & Media | NaN |
| **3** | 1970 | 1 | 0 | Greece | Attica | Western Europe | 37.997490 | 23.762728 | 1 | Bombing/Explosion | Athens | Government (Diplomatic) | NaN |
| **4** | 1970 | 1 | 0 | Japan | Fukouka | East Asia | 33.580412 | 130.396361 | 1 | Facility/Infrastructure Attack | Fukouka | Government (Diplomatic) | NaN |

## removing success as we do not need it as all values have success value =1

In [9]:

```
dff.drop('success',axis=1,inplace=True)
```

In [10]:

```
dff.dtypes
```

Out[10]:

```
iyear                  int64
imonth                 int64
iday                   int64
country_txt           object
provstate             object
region_txt            object
latitude             float64
longitude            float64
attacktype1_txt       object
city                  object
targtype1_txt         object
motive                object
gname                 object
weaptype1_txt         object
dtype: object
```

In [11]:

```
dff.describe()
```

Out[11]:

| | iyear | imonth | iday | latitude | longitude |
|---|---|---|---|---|---|
| **count** | 161632.000000 | 161632.000000 | 161632.000000 | 157309.000000 | 1.573080e+05 |
| **mean** | 2002.251472 | 6.463881 | 15.468997 | 23.012393 | -5.210354e+02 |
| **std** | 13.247559 | 3.385112 | 8.814507 | 18.678939 | 2.173006e+05 |
| **min** | 1970.000000 | 0.000000 | 0.000000 | -53.154613 | -8.618590e+07 |
| **25%** | 1990.000000 | 4.000000 | 8.000000 | 10.686589 | 3.594444e+00 |
| **50%** | 2008.000000 | 6.000000 | 15.000000 | 31.200657 | 4.314357e+01 |
| **75%** | 2014.000000 | 9.000000 | 23.000000 | 34.535939 | 6.844713e+01 |

| | iyear | imonth | iday | latitude | longitude |
|---|---|---|---|---|---|
| **max** | 2017.000000 | 12.000000 | 31.000000 | 74.633553 | 1.793667e+02 |

In [12]:
```
dff.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 161632 entries, 0 to 181688
Data columns (total 14 columns):
 #   Column           Non-Null Count   Dtype  
---  ------           --------------   -----  
 0   iyear            161632 non-null  int64  
 1   imonth           161632 non-null  int64  
 2   iday             161632 non-null  int64  
 3   country_txt      161632 non-null  object 
 4   provstate        161245 non-null  object 
 5   region_txt       161632 non-null  object 
 6   latitude         157309 non-null  float64
 7   longitude        157308 non-null  float64
 8   attacktype1_txt  161632 non-null  object 
 9   city             161226 non-null  object 
 10  targtype1_txt    161632 non-null  object 
 11  motive           46402 non-null   object 
 12  gname            161632 non-null  object 
 13  weaptype1_txt    161632 non-null  object 
dtypes: float64(2), int64(3), object(9)
memory usage: 18.5+ MB
```

## country wise damage/attacks

In [28]:
```
country_plot=dff.country_txt.value_counts()
country_plot.to_frame().head(20)
```

Out[28]:

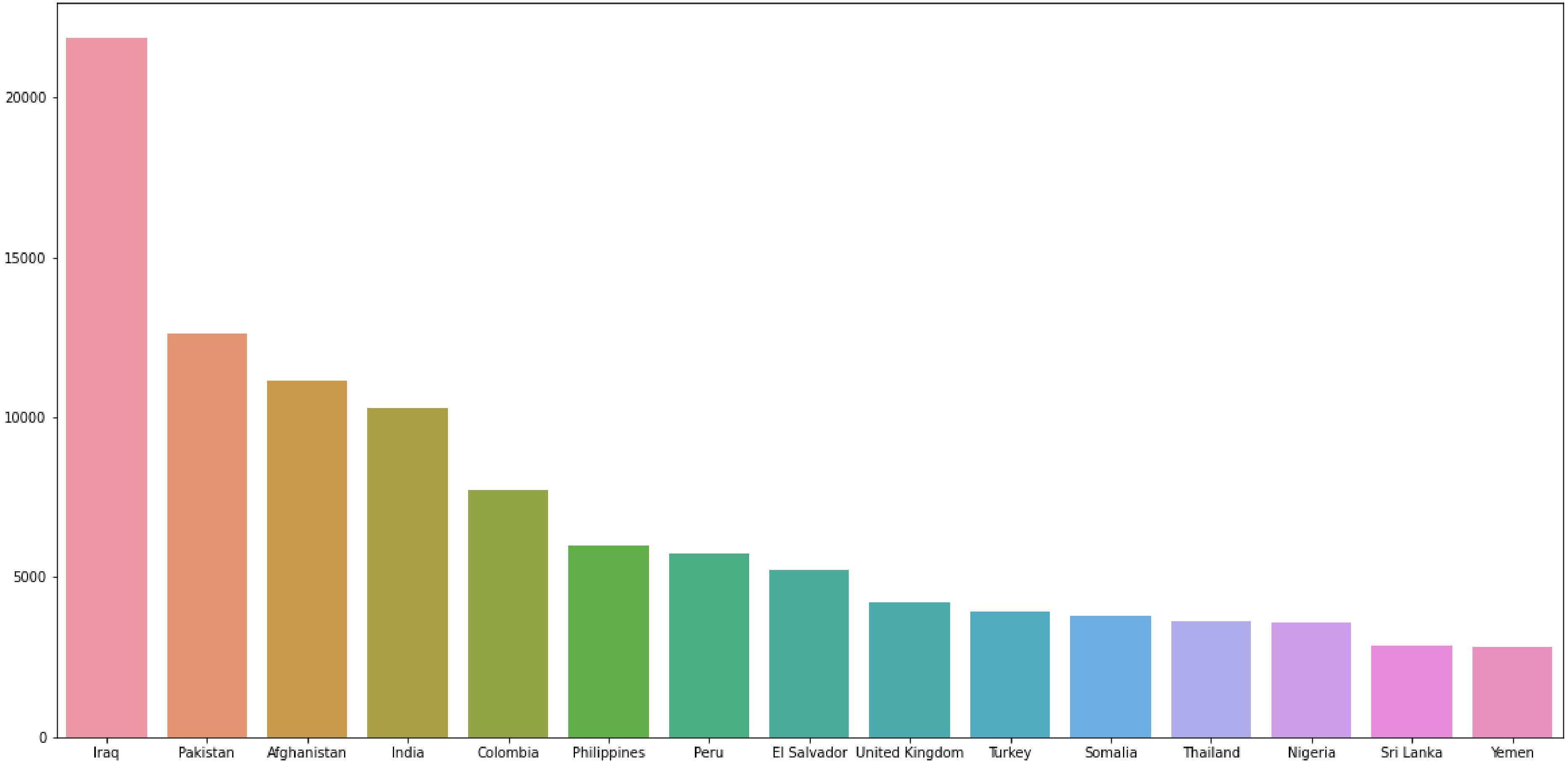
| | country_txt |
|---|---|
| **Iraq** | 21861 |
| **Pakistan** | 12600 |

| | country_txt |
|---|---|
| **Afghanistan** | 11141 |
| **India** | 10280 |
| **Colombia** | 7712 |
| **Philippines** | 5975 |
| **Peru** | 5755 |
| **El Salvador** | 5227 |
| **United Kingdom** | 4206 |
| **Turkey** | 3909 |
| **Somalia** | 3804 |
| **Thailand** | 3626 |
| **Nigeria** | 3593 |
| **Sri Lanka** | 2849 |
| **Yemen** | 2837 |
| **Spain** | 2818 |
| **Algeria** | 2561 |
| **France** | 2481 |
| **United States** | 2340 |
| **Chile** | 2221 |

In [14]:

```
plt.figure(figsize=(20,10))
sns.barplot(dff['country_txt'].value_counts()[:15].index,dff['country_txt'].value_counts()[:15].values)
```

Out[14]: <AxesSubplot:>




## region wise attacks counts

In [15]:

```
dff.region_txt.value_counts().to_frame()
```

Out[15]:

| | region_txt |
|---|---|
| **Middle East & North Africa** | 44319 |
| **South Asia** | 39369 |
| **South America** | 17620 |
| **Sub-Saharan Africa** | 16277 |

| | region_txt |
|---|---|
| **Western Europe** | 14161 |
| **Southeast Asia** | 11151 |
| **Central America & Caribbean** | 9979 |
| **Eastern Europe** | 4437 |
| **North America** | 2894 |
| **East Asia** | 680 |
| **Central Asia** | 505 |
| **Australasia & Oceania** | 240 |

# types of attacks and its counts

In [16]:

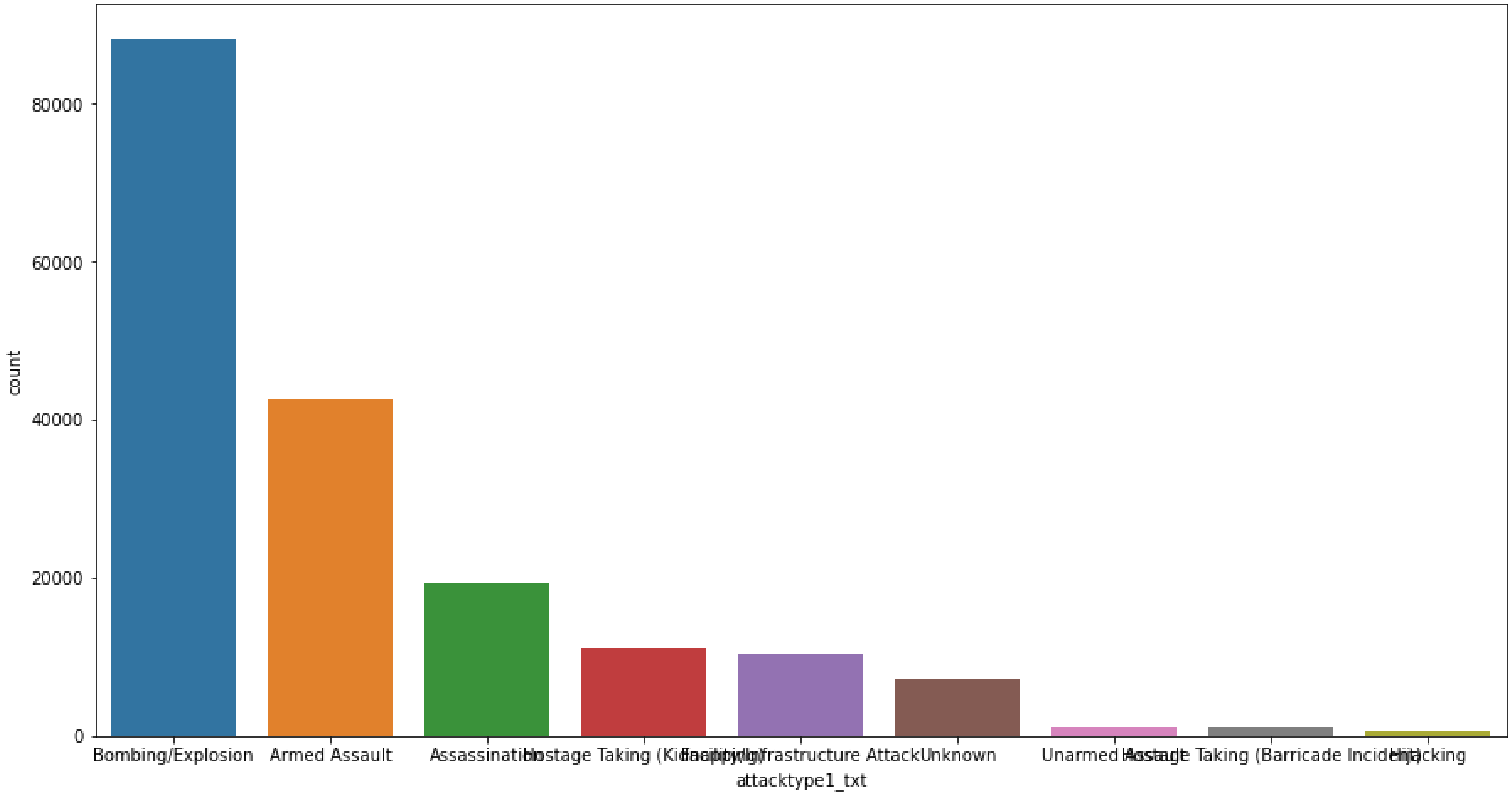
```
dff.attacktype1_txt.value_counts().to_frame()
```

Out[16]:

| | attacktype1_txt |
|---|---|
| **Bombing/Explosion** | 77530 |
| **Armed Assault** | 40345 |
| **Assassination** | 14615 |
| **Hostage Taking (Kidnapping)** | 10907 |
| **Facility/Infrastructure Attack** | 9779 |
| **Unknown** | 6015 |
| **Hostage Taking (Barricade Incident)** | 983 |
| **Unarmed Assault** | 879 |
| **Hijacking** | 579 |

In [17]:
```python
plt.figure(figsize=(15,8))
sns.countplot('attacktype1_txt',data=df,order=df['attacktype1_txt'].value_counts().index)
```

Out[17]: <AxesSubplot:xlabel='attacktype1_txt', ylabel='count'>




In [18]:
```python
dff.city.value_counts().to_frame().head(20)
```

Out[18]:

| | city |
|---|---|
| **Unknown** | 8705 |

| | city |
|---|---|
| **Baghdad** | 7226 |
| **Karachi** | 2428 |
| **Lima** | 2176 |
| **Mosul** | 1902 |
| **Belfast** | 1797 |
| **Santiago** | 1509 |
| **San Salvador** | 1495 |
| **Mogadishu** | 1444 |
| **Istanbul** | 935 |
| **Athens** | 897 |
| **Bogota** | 892 |
| **Kirkuk** | 844 |
| **Beirut** | 793 |
| **Medellin** | 782 |
| **Benghazi** | 756 |
| **Quetta** | 716 |
| **Baqubah** | 694 |
| **Guatemala City** | 680 |
| **Peshawar** | 678 |

## Gang names including unknown gangs

In [19]:

```
dff.gname.value_counts().to_frame()
```

Out[19]:

| | gname |
|---|---|
| **Unknown** | 71748 |
| **Taliban** | 6680 |
| **Islamic State of Iraq and the Levant (ISIL)** | 4759 |
| **Shining Path (SL)** | 4337 |
| **Farabundo Marti National Liberation Front (FMLN)** | 3317 |
| **...** | ... |
| **Association of Mobil Spill Affected Communities (AMSAC)** | 1 |
| **New Revolutionary Alternative (NRA)** | 1 |
| **Pemuda Pancasila** | 1 |
| **National Democratic Alliance of Sudan** | 1 |
| **MANO-D** | 1 |

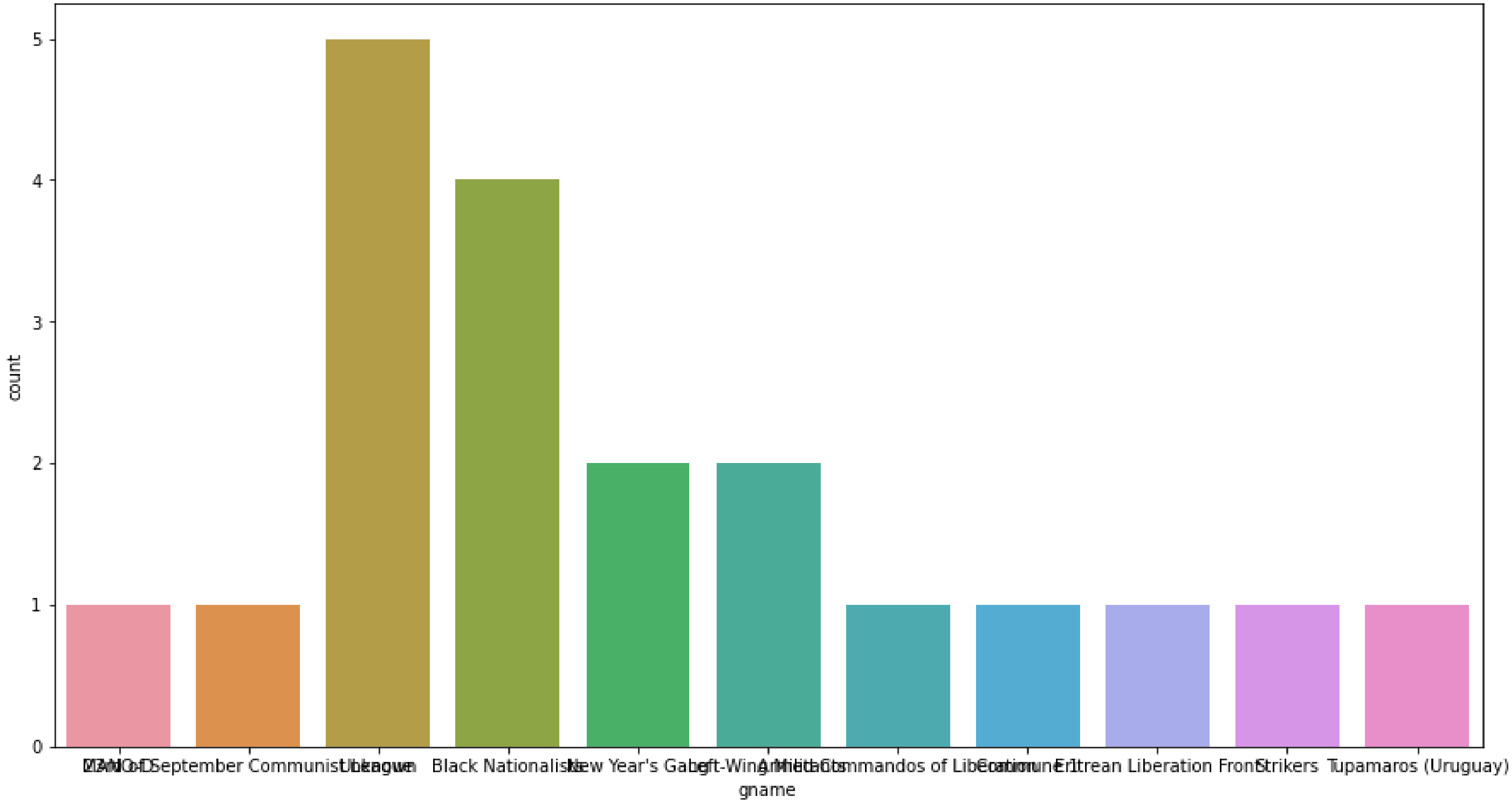
3334 rows × 1 columns

In [20]:

```
plt.figure(figsize=(15,8))
sns.countplot(data = dff[:20], x = 'gname')
```

Out[20]:

```
<AxesSubplot:xlabel='gname', ylabel='count'>
```




## yearly attacks

In [21]:
```
dff.iyear.value_counts().to_frame()
```

Out[21]:

| | iyear |
|---|---|
| **2014** | 15015 |
| **2015** | 12676 |

| | iyear |
|---|---|
| **2016** | 10975 |
| **2013** | 10484 |
| **2017** | 8652 |
| **2012** | 7600 |
| **2011** | 4606 |
| **1992** | 4560 |
| **2009** | 4430 |
| **2008** | 4402 |
| **2010** | 4385 |
| **1991** | 4299 |
| **1989** | 4008 |
| **1990** | 3583 |
| **1988** | 3440 |
| **1984** | 3282 |
| **2007** | 3123 |
| **1994** | 3069 |
| **1997** | 2966 |
| **1987** | 2933 |
| **1995** | 2794 |
| **1996** | 2770 |
| **1985** | 2727 |
| **1986** | 2670 |
| **1983** | 2660 |

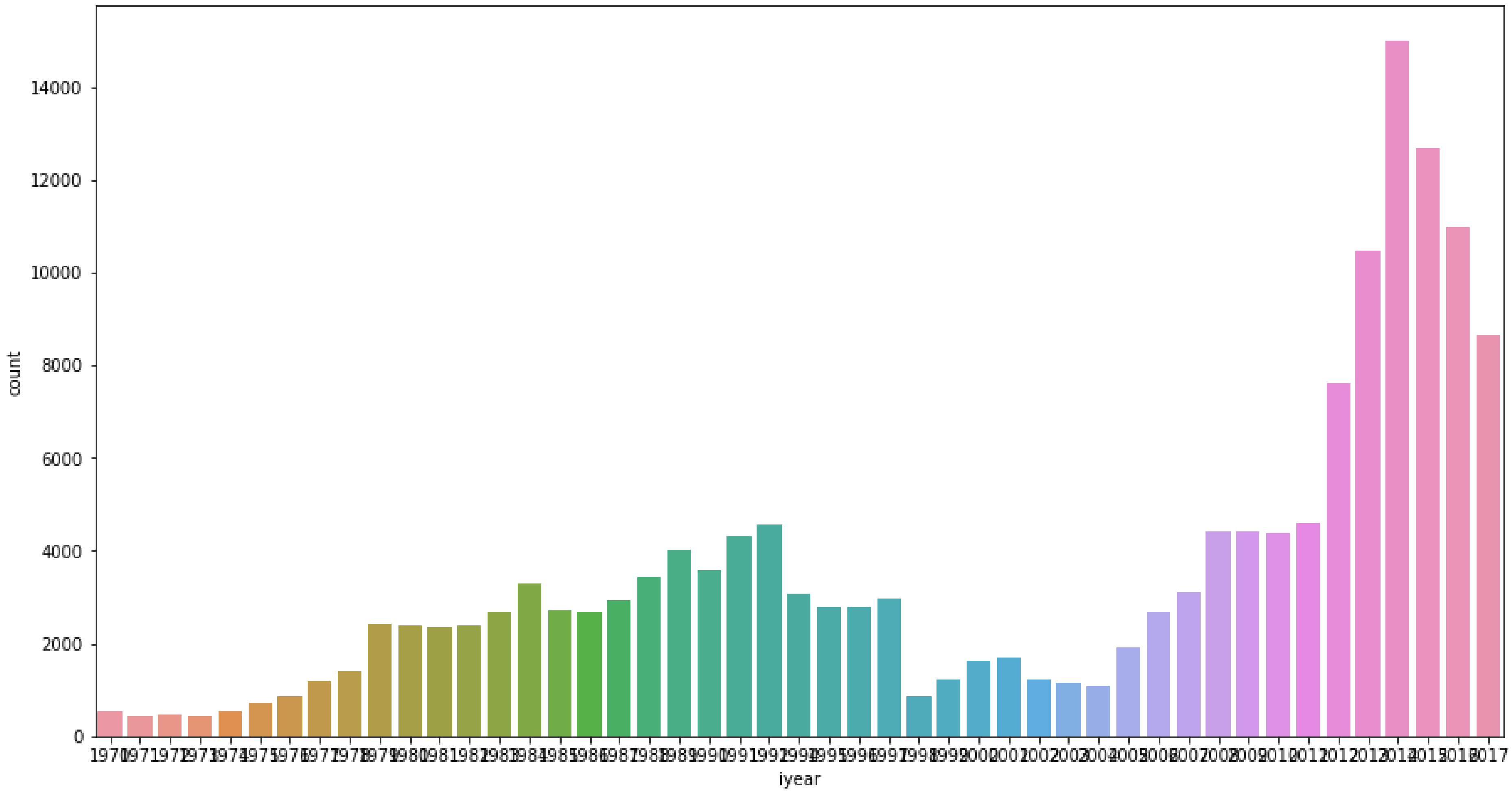
| | iyear |
|---|---|
| **2006** | 2660 |
| **1979** | 2408 |
| **1980** | 2387 |
| **1982** | 2373 |
| **1981** | 2354 |
| **2005** | 1910 |
| **2001** | 1689 |
| **2000** | 1637 |
| **1978** | 1411 |
| **1999** | 1237 |
| **2002** | 1213 |
| **1977** | 1191 |
| **2003** | 1149 |
| **2004** | 1080 |
| **1976** | 861 |
| **1998** | 859 |
| **1975** | 705 |
| **1970** | 549 |
| **1974** | 545 |
| **1972** | 452 |
| **1973** | 433 |
| **1971** | 420 |

In [22]:

```
plt.figure(figsize=(15,8))
sns.countplot(data = dff, x = 'iyear')
```

Out[22]:

```
&lt;AxesSubplot:xlabel='iyear', ylabel='count'&gt;
```




In [23]:

```
dff.nunique()
```

Out[23]:

```
iyear              47
imonth             13
iday               32
country_txt       202
```

```
provstate           2742
region_txt            12
latitude           43347
longitude          43099
attacktype1_txt        9
city               33923
targtype1_txt         22
motive             13298
gname               3334
weaptype1_txt         12
dtype: int64
```

In [24]:

```
dff.describe()
```

Out[24]:

| | iyear | imonth | iday | latitude | longitude |
|---|---|---|---|---|---|
| **count** | 161632.000000 | 161632.000000 | 161632.000000 | 157309.000000 | 1.573080e+05 |
| **mean** | 2002.251472 | 6.463881 | 15.468997 | 23.012393 | -5.210354e+02 |
| **std** | 13.247559 | 3.385112 | 8.814507 | 18.678939 | 2.173006e+05 |
| **min** | 1970.000000 | 0.000000 | 0.000000 | -53.154613 | -8.618590e+07 |
| **25%** | 1990.000000 | 4.000000 | 8.000000 | 10.686589 | 3.594444e+00 |
| **50%** | 2008.000000 | 6.000000 | 15.000000 | 31.200657 | 4.314357e+01 |
| **75%** | 2014.000000 | 9.000000 | 23.000000 | 34.535939 | 6.844713e+01 |
| **max** | 2017.000000 | 12.000000 | 31.000000 | 74.633553 | 1.793667e+02 |

In [25]:

```
dff.corr()
```

Out[25]:

| | iyear | imonth | iday | latitude | longitude |
|---|---|---|---|---|---|
| **iyear** | 1.000000 | -0.001319 | 0.015207 | 0.174379 | 0.004070 |
| **imonth** | -0.001319 | 1.000000 | 0.003928 | -0.016112 | -0.004124 |
| **iday** | 0.015207 | 0.003928 | 1.000000 | 0.003795 | -0.002435 |

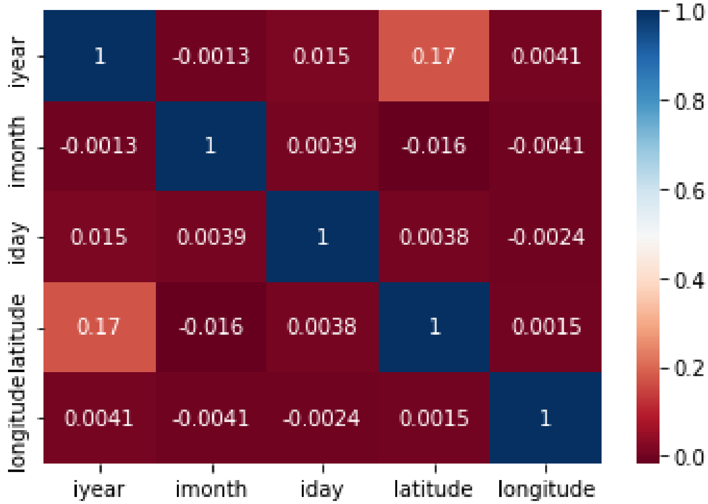
| | iyear | imonth | iday | latitude | longitude |
|---|---|---|---|---|---|
| **latitude** | 0.174379 | -0.016112 | 0.003795 | 1.000000 | 0.001475 |
| **longitude** | 0.004070 | -0.004124 | -0.002435 | 0.001475 | 1.000000 |

## Heatmap relating correlation of all terms

In [26]:

```
sns.heatmap(dff.corr(),annot=True,cmap='RdBu')
```

Out[26]: <AxesSubplot:>




## Thank You